

4E-5

用例検索翻訳支援システム

中村直人

(株)富士通研究所

1. まえがき

翻訳を行う際に、用例を参照して訳語や表現を選択することが重要であると言われている。これは、単語や表現が使われている適切な実例等を、翻訳者は用例から得ることができるからであろう。今日、種々の翻訳支援/電子化辞書システムが研究・開発されている^(1,2,3)が、まだ用例を参照するために検索する機能を持ったものは少ない。

本稿で紹介する用例検索翻訳支援システム(試作システム)は、翻訳原文中の部分、時には原文一文、を入力して用例を検索するシステムである。本システムの方式は、単に見出し語で検索する通常の辞書とは異なる検索法であるばかりでなく、ワープロ等で翻訳原文から切り出した文字列をそのまま入力として使用できるという利点がある。本システムでは、自立語によるインデックス検索という比較的低コストの処理法を基本としながらも、入力が含んでいる情報を活かした検索を提案している。

本稿では、システムの特徴をその構成と共に紹介して、最後に小規模な評価の結果を示してシステムの有効性とユーザの振る舞いについて述べる。

2. 用例検索の仕様

ここで検討する用例検索の設定は、日英の対訳を用例として扱い、入力は日本語とする。システムに検索のための入力(以下、語や句である場合も含めて「入力文」と呼ぶ)を与えると、入力文を解析した結果を元に入力文との関連の高い用例を検索して表示する(本システム例、図1)。ユーザは表示された用例を、検索目的に合致するかどうかで判定して選択する。

我々は、次のような機能を持った用例検索システムが翻訳支援に有効であると考えた。

- ・単語だけでなく自由な形式の入力文で検索できる
- ・検索結果に応じた検索条件の変更が簡単に行える
- ・ユーザが登録可能な辞書の構成法を採る

一般に、用例を検索する手法として、通常の見出し語を用いた辞書、KWIC、キーワード設定によるデータ検索などが使われている。しかしこれらは上記の仕様を十分に満たしていないので、これらと異なる枠組みでの用例検索システムを作成した。

以下では、用例検索に上記の仕様を要求する背景について述べる。

2.1. 自由な形式の入力文による検索

複数の語を含む入力文を受け付ければ、検索する内容を代表する見出し語を選ばずに検索することができる。例えば「頭をひねる」という表現の検索においてこの表現をそのまま指定することができれば、「頭」や「ひねる」から一語を見出し語として選択する必要がない。

入力文に句や文を許せば、翻訳原文が既に文書データとして入力されている場合に、原文から切り出した部分を入力文として使うことができる。

ただし、入力文が単語以上の情報を持つと、検索機構には柔軟性が要求される。なぜなら、一般に、複数の用例が検索され、これらの用例間で入力文との相関にばらつきが生じるので、この相関の度合いにより適当に選択して検索結果とする必要がある。

入力文を元に柔軟な検索を行うための手法をとして、ある規則に従って入力文を一般化する手法⁽⁴⁾や意味解析とソーラスの利用などがある。このような内部の規則・知識に従って入力文を処理する方式では、検索条件が内部的に変更されるため、かえってユーザの意図した通りに検索されない場合がある。例えば、修飾語を補助的なものとして削除する一般化を採用したシステムの例を挙げると、「質の良い例文」を入力した場合、検索の重きが「質の良い～」にあっても、「質の良い」が一般化によって削られ「例文」が残ってしまい良い結果は期待できないのである。

2.2. 検索条件変更の容易さ

用例検索の結果は入力文と用例辞書の内容によるので、用例が多数検索されたり、全く検索されなかったりする場合がある。用例検索ではデータ検索の専門家ではないユーザを対象としているので、検索条件の変更が容易であることが必要である。

本システムでは、入力文の個々の自立語について要・不要を指定することによって検索条件を変更できる。例えば、「ボートで川を下る」を入力して多数の用例が検索されてしまった場合に、「ボート」「川」「下る」の中からとりわけ検索したい語を幾つか選択して、検索条件を調整することができる。

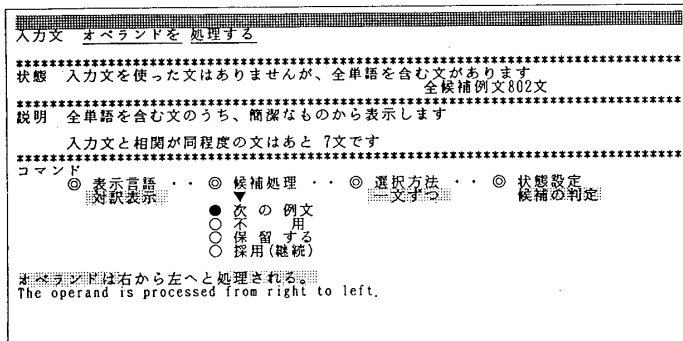


図1 システムの動作例

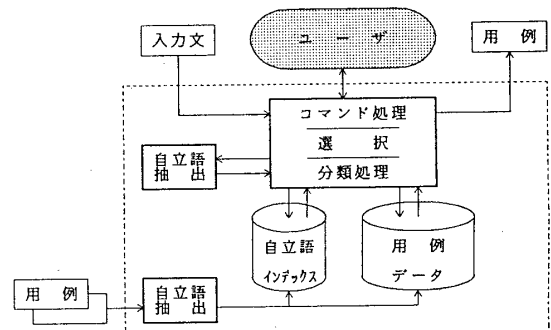


図2 システムの構成

Translation support by retrieving bilingual texts

Naohito NAKAMURA

FUJITSU Laboratories, LTD.

2.3. ユーザが登録可能な辞書

用例辞書の使い方は、あらかじめ辞書を作成しておきユーザは専ら検索のみ行う場合と、ユーザも必要に応じて用例の追加を行う場合がある。後者の場合、ユーザ自身が用例を登録するので、通常の辞書に情報を付加する場合のような系統立った作業なしに自動登録できる方式が便利である。

容易な用例登録を実現することにより、豊富な用例を低コストで蓄積することができ、かつユーザによる追加も可能になる。

3. 辞書の構成・検索手法

3.1. 辞書の構成法

本システムでは用例を登録時に自立語解析して、文中の自立語（以下これを「構成語」と呼ぶ）をインデックスとして辞書に登録している（図2）。この方式により、用例の辞書中での位置づけ・分類等の作業なしに、用例を追加して行くことが可能になる。

3.2. 検索

入力文も登録と同様な手法で自立語解析し、構成語を得る。そして、この構成語を含む用例を検索し、その結果を母集合として、次の分類条件に従って部分集合に分類する。入力文を「進歩が著しい」として例を付記する。

集合A … 入力文が用例の中に現れる（入力文による文字列検索に相当）。要素例：「機械翻訳の進歩が著しい」

集合B … 入力文の全構成語を用例が含んでいる（構成語の全数マッチ）。要素例：「著しく進歩した技術」

集合C … 構成語を幾つか含んでいる（母集合）。要素例：「技術の進歩と製品の価格」

これらの集合の関係は、次のようになる。

$$\text{集合A} \subseteq \text{集合B} \subseteq \text{集合C} \dots\dots (1)$$

式(1)の性質を利用して、本システムでは文字列検索結果に相当する集合Aを作成している。つまり、まず集合Cから集合Bを選び出し、さらに集合Bから集合Aを選び出す、という手順で行う。

3.3. ユーザへの提示

検索結果のユーザへの提示は、(3.2.節の)集合Aから集合Cの順で（重複要素は一回のみ）、しかも、各集合内では以下に述べる整理を施して行う。この整理の目的は、構成語が局所的に出現して、文の長さが短い（単語数の少ない）文を先にユーザに提示することである。

- ① 構成語がその他の語を挟まずに連続している。
- ② 用例中の自立語の個数に対する構成語の個数の比が高い。
- ③ 含んでいる構成語の個数が多い。

4. ユーザの振る舞い

本システムを十数名のユーザに試用させた結果として得たユーザの振る舞い、特に、句や文を入力して検索するという本システムの特徴的な部分についてまとめた。

評価は、あらかじめ辞書データを登録しておき、ユーザには文書を翻訳する課題を与え、検索のみを行わせた。評価用システムは単独で実現されており、ワープロ等との連携は行っていない。ユーザは文書から必要な箇所を適当に手で入力して検索した。試用したユーザは、翻訳家3名とソフトウェア研究者10名である。この評価での入力件数はのべ約百件であった。

4.1. 入力文の文節数

ユーザが原文中で検索対象とした箇所と、それに対応する実際に入力文の長さを文節数で比較した（図3）。ユーザが検索の対象にした原文中の箇所の半数が多文節であり、これをうけて、入力文も多文節が多く用いられている。この評価において、通常の見出し語式辞書の機能を果たすシステムを併用していれば、本システムはさらに多文節での検索に専用化されたであろう。

4.2. 入力文生成のタイプ

原文と入力文の比較を行った（図4）。本システムの特徴を反映して、入力文の7割程度が原文のままである。

入力文が多文節である場合で原文と異なる入力をしてい

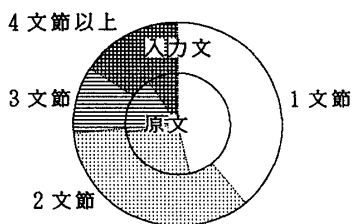


図3 原文と入力文の文節数

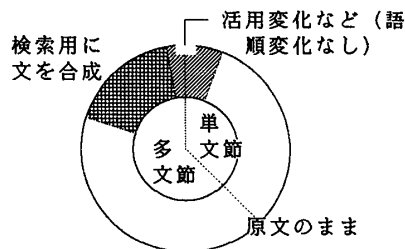


図4 入力文の生成タイプ

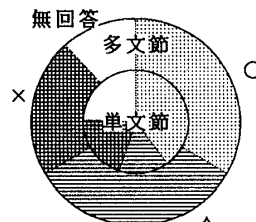


図5 検索結果の評価

る例では、通常の辞書検索のような活用語の終止形化、文構造の変形がある。文構造の変形は多文節入力に特徴的なものである。この例としては、原文「…は、あらかじめ…サブルーチンにより定義されていなければならない。」を、一部の単語を省いて「あらかじめ定義されている」と入力しているものなどがある。

4.3. 検索結果に対する評価

個々の検索ごとに検索結果が期待通りのものであったかの評価をユーザに求めた（図5）。

入力文が多文節の場合は、単文節の場合に比べて△の評価が多い。これは、単文節（単語）で検索した場合に比べて、多文節で引いた場合は期待する検索結果があいまいなのであろう。なお、評価は検索機構自身の性能と辞書データの質の両者によって決まるので、この判定からシステム自身の評価は行い難い。

5. おわりに

入力文と用例との相関を、両者が共有する自立語で見積もった検索システムを試作した。簡単な評価の結果、翻訳中に生じる用例の検索要求には句や文で表されるようなものも多いことが分かった。

今後は、実際の翻訳の場で、入力文とその検索結果に関するユーザの要求を調査し、用例検索により相応しい検索機能を研究してゆく必要があるだろう。

〔参考文献〕

- (1) 信国 他：日英機械翻訳のための和文推敲支援システム(1)、情処学会、第35回全国大会論文集、3S-8 (Sep. 1987)
- (2) Isoda, M., Aiso, H. : Model for Lexical Knowledge Base, Proceedings of 11th International Conference on Computational Linguistics (Coling '86), pp. 451-453 (Aug. 1986)
- (3) 隅田、堤：構文の照合による柔軟なテキスト検索機能を備えた翻訳支援システム、情処学会、第37回全国大会論文集、4B-6 (Sep. 1988)